Ahmed Mabrouk , Abdelrahman Saed , Yasser Ahmed , Youssef Diaa , Alaa Ahmed

Under supervision : Dr/ Osama Hefny

AI Engineer Lecturer Assistant at Faculty of Computers and Artificial intelligence Master's degree, Computer Science Master's degree Beni-suef university

Research Paper on Sentiment Analysis of E-Commerce Customer Reviews

Abstract

This research paper explores the implementation of sentiment analysis techniques in the e-commerce sector, specifically targeting customer reviews for women's clothing. By leveraging various supervised machine learning models, the study aims to classify sentiments expressed in product reviews, thereby providing e-commerce businesses with valuable insights into consumer preferences and enhancing customer relationship management strategies.

Introduction

Sentiment analysis, also referred to as opinion mining, is a pivotal technique in natural language processing (NLP) that assesses textual data to determine sentiment polarity—positive, negative, or neutral. In the e-commerce landscape, understanding customer feedback is crucial for making informed business decisions. This study addresses the challenge of analyzing unstructured customer reviews to extract actionable insights, which can significantly influence purchasing decisions and brand loyalty.

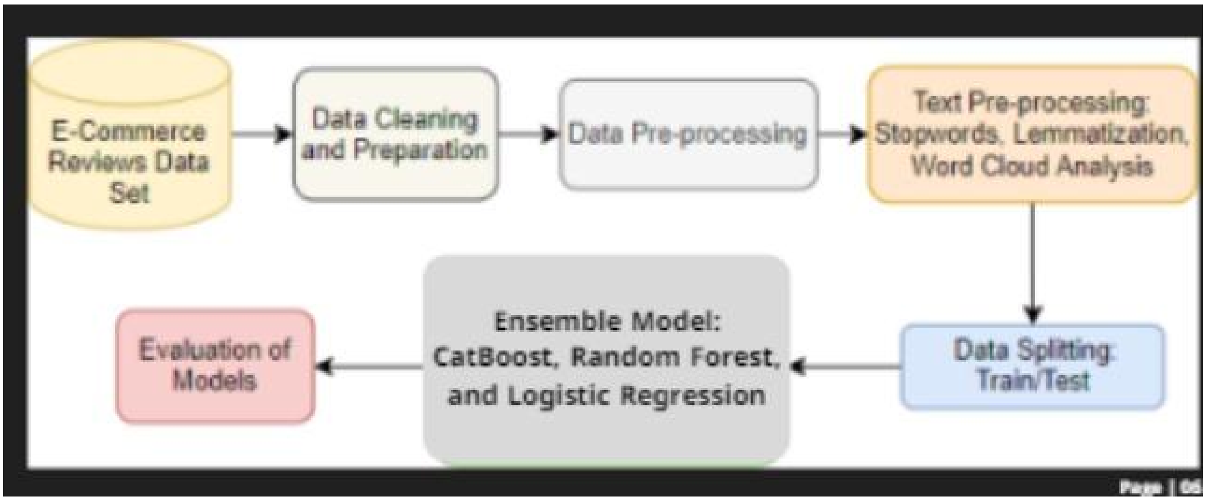
Ensemble Model:

(CatBoost, Random Forest, and Logistic Regression)

Overview:
The ensemble model integrates the predictive power of CatBoost and Random Forest, combining them through Logistic Regression as the final estimator. This hybrid approach leverages the strengths of both tree-based models and linear classifiers to enhance overall predictive performance and robustness.

Methodology

The methodology is structured into three comprehensive steps: data collection and pre-processing, feature extraction, and model training and evaluation.



1- Data set description

The data set "E-Commerce Customer Reviews" used in our experiments is composed of the following variables: "Clothing ID", "Age", "Title", "Rating", "Review Text", "Recommended IND", "Positive Feedback Count", "Division Name", "Class Name", and "Department Name" (Fig.2).

	Unnamed: 0	Clothing ID	Age	Title	Review Text	Rating	Recommended IND	Positive Feedback Count	Division Name	Department Name	Class Name
0	0	767	33	NaN	Absolutely wonderful - silky and sexy and comf...	4	1	0	Initmates	Intimate	Intimates
1	1	1080	34	NaN	Love this dress! it's sooo pretty. i happene...	5	1	4	General	Dresses	Dresses
2	2	1077	60	Some major design flaws	I had such high hopes for this dress and reall...	3	0	0	General	Dresses	Dresses
3	3	1049	50	My favorite buy!	I love, love, love this jumpsuit. it's fun, fl...	5	1	0	General Petite	Bottoms	Pants
4	4	847	47	Flattering shirt	This shirt is very flattering to all due to th...	5	1	6	General	Tops	Blouses
...	...	...	...	...	...	...	...	...	...	...	...
23481	23481	1104	34	Great dress for many occasions	I was very happy to snag this dress at such a ...	5	1	0	General Petite	Dresses	Dresses
23482	23482	862	48	Wish it was made of cotton	It reminds me of maternity clothes. soft, stre...	3	1	0	General Petite	Tops	Knits
23483	23483	1104	31	Cute, but see through	This fit well, but the top was very see throug...	3	0	1	General Petite	Dresses	Dresses
23484	23484	1084	28	Very cute dress, perfect for summer parties an...	I bought this dress for a wedding i have this ...	3	1	2	General	Dresses	Dresses
23485	23485	1104	52	Please make more like this one!	This dress in a lovely platinum is feminine an...	5	1	22	General Petite	Dresses	Dresses

23486 rows × 11 columns

Fig. 2: E-commerce reviews data set

2- **Data pre-processing and exploration**

In this step, we removed the missing data from each variable. Group the product title variable with the comment text. Then we showed the distribution of the target variable "Recommend_IND" (Figure 3). We can notice that our data is not balanced because the majority of the customers recommend the purchased products (more than 80%). Next create a new variable that represents the length of each comment "Text_Length". And analyze the relationship between the text length variable and the target variable (Fig. 4). Then visualize the correlation between the target variable and the "Rating" variable (Fig.5, 6). Common Python libraries used to perform pre-processing tasks including "NLTK" (Natural Language Toolkit) and "RE" (Regular Expression) [14].

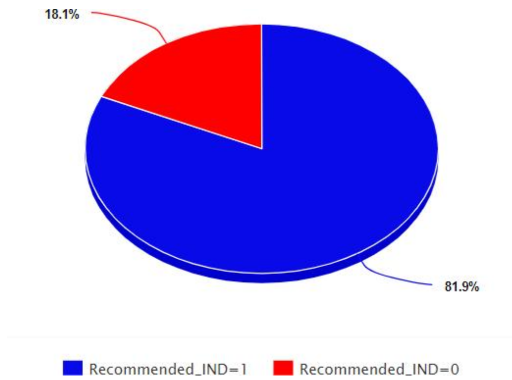


Fig. 3: The pie chart's percentages for recommended and unrecommended products

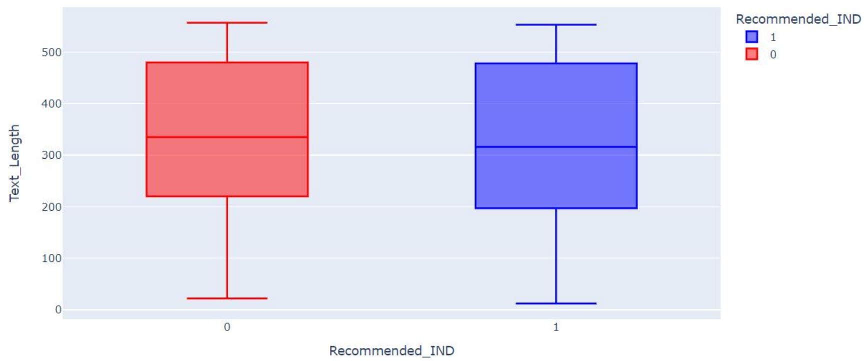


Fig. 4: The target variable vs. text length box plot

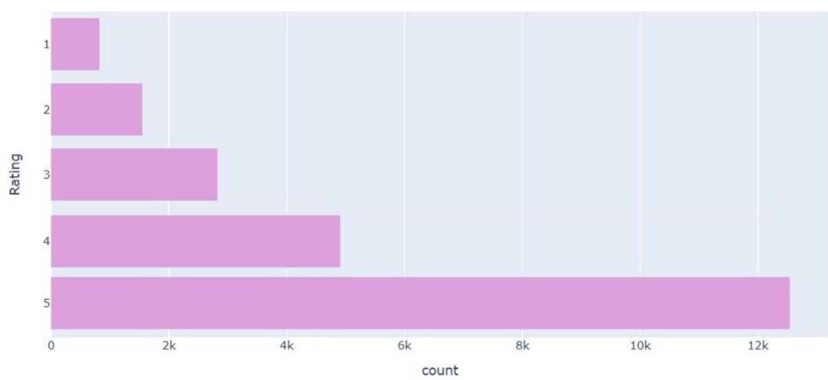


Fig. 5: product rating bar plot

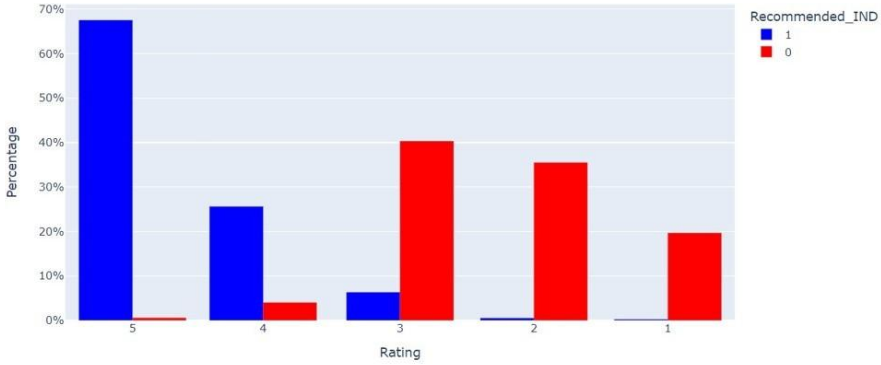


Fig. 6: The target variable vs. product rating

Polarity analysis

Polarity is a floating value that lies in the interval [-1,1] where 1 signifies positive feedbacks and -1 a negative one [15]. The distribution of the polarity score in the customers' reviews is shown in Figure 7. Where the majority of the comments are situated on the positive side of the graph [0,1].

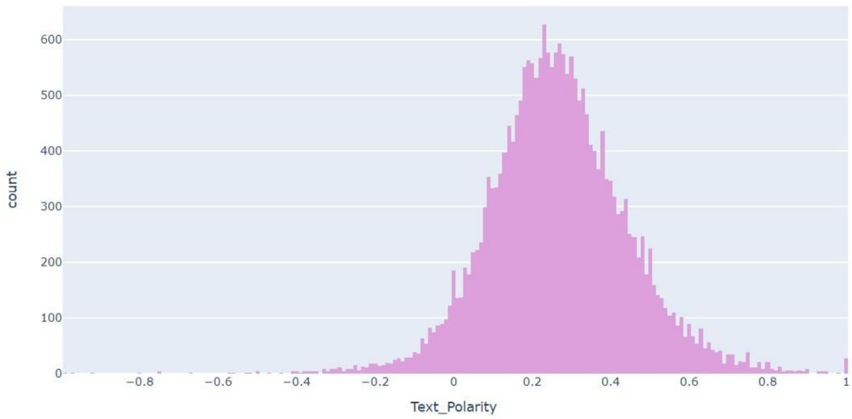


Fig. 7: Reviews text polarity bar plot

Text pre-processing

In this phase, punctuation is removed (!"#$%&\'()*+,-./:;<=>?@[\\]^_`{|}~), and all text in the comments is converted to lowercase. Based on the values of the Text_Polarity variable two different sentiments can be derived. Positive sentiment is present if Text_Polarity is larger than zero. Conversely, a negative sentiment is present if Text_Polarity is less than zero (Fig. 8).

	Rating	Class_Name	Recommended_IND	Text	Text_Length	Text_Polarity	Sentiment
0	4	Intimates	1	absolutely wonderful silky and sexy and comf...	54	0.633	Positive
1	5	Dresses	1	love this dress its sooo pretty i happened ...	304	0.340	Positive
2	3	Dresses	0	some major design flaws i had such high hopes ...	524	0.073	Positive
3	5	Pants	1	my favorite buy i love love love this jumpsuit...	141	0.561	Positive
4	5	Blouses	1	flattering shirt this shirt is very flattering...	209	0.513	Positive
5	2	Dresses	0	not for the very petite i love tracy reese dre...	512	0.181	Positive
6	5	Knits	1	cagrcoal shimmer fun i aded this in my basket ...	517	0.158	Positive
7	4	Knits	1	shimmer surprisingly goes with lots i ordered ...	519	0.230	Positive
8	5	Dresses	1	flattering i love this dress i usually get an ...	177	0.003	Positive
9	5	Dresses	1	such a fun dress im 55 and 125 lbs i ordered t...	378	0.202	Positive
10	3	Dresses	0	dress looks like its made of cheap material dr...	381	-0.047	Negative
11	5	Dresses	1	this dress is perfection so pretty and flatte...	52	0.250	Positive

Fig. 8: Reviews text polarity and sentiment data set

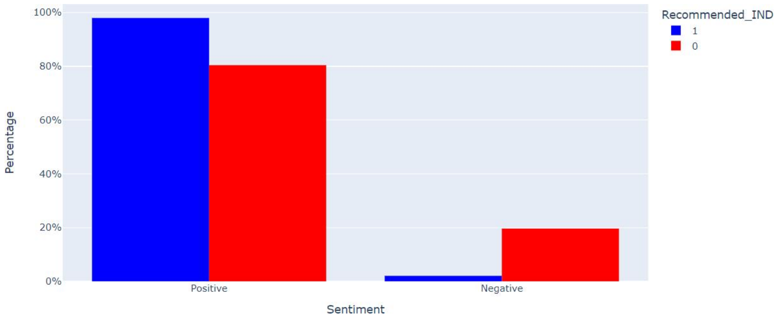


Fig. 9: Percentage of sentiments in relation to the target variable bar plot

Stop-words, stemming, and lemmatization

This step consists of removing regular expressions and stop words, stemming, and lemmatizing the text of the reviews. Stemming is the practice of removing the final few characters from a word, which frequently results in inaccurate spelling and meanings. By taking context into account, lemmatization reduces a term to its logical base form, or lemma [16].

Example:

- Original : changing, arrived
- Stemming : chang, arriv
- Lemmatization : change, arrive

Word-cloud analysis

Finally, common words have been removed that do not affect the prediction of negative or positive reviews. The frequency of words is shown graphically in word-clouds. The size of the word in the created graphic increases in proportion to how frequently the keyword appears in the analyzed text reviews [17]. The positive word-cloud is represented in Figure 10 and the negative word-cloud in Figure 11. The final data set containing the updated version of the reviews and the target variable is shown in Figure 12.

Fig. 10: Positive word-cloud

Fig. 11: Negative word-cloud

	Updated_Review_Text	Recommended_IND
0	absolutely wonderful silky sexy comfortable	1
1	love sooo pretty happen find store glad bc nev...	1
2	major design flaws high hope really wanted wor...	0
3	favorite buy love love love jumpsuit fun flirt...	1
4	flatter flatter due adjustable front tie perfe...	1
5	petite love tracy reese petite 5 foot tall usu...	0
6	cagrcoal shimmer fun aded basket hte last mint...	1
7	shimmer surprisingly go lot carbon store pick ...	1
8	flatter love usually get xs run little snug bu...	1
9	fun 55 125 lb petite make sure length wasnt lo...	1

Fig. 12: Review text and the target variable data set

1 Experimental Results and Discussion

Confusion matrix

The outcomes of predictions on a classification task are summarized in a confusion matrix. Correct and incorrect predictions are highlighted and divided into classes. The result of predictions is compared with the real values. The representation of the confusion matrices is shown in Figure 13. True positives and false positives are denoted TP and FP, while false negatives and true negatives are denoted FN and TN. Where:

- TP: The number of clients who have recommended a product (class 1), and whom the predictive model correctly predicted.
- TN: The number of clients who did not recommended a product (class 0), and that the predictive model correctly predicted.
- FP: The number of customers who did not recommended a product (class 0), but that the predictive algorithm identified as class 1.
- FN: The number of customers who have recommended a product (class 1), but that the predictive model identified as class 0.

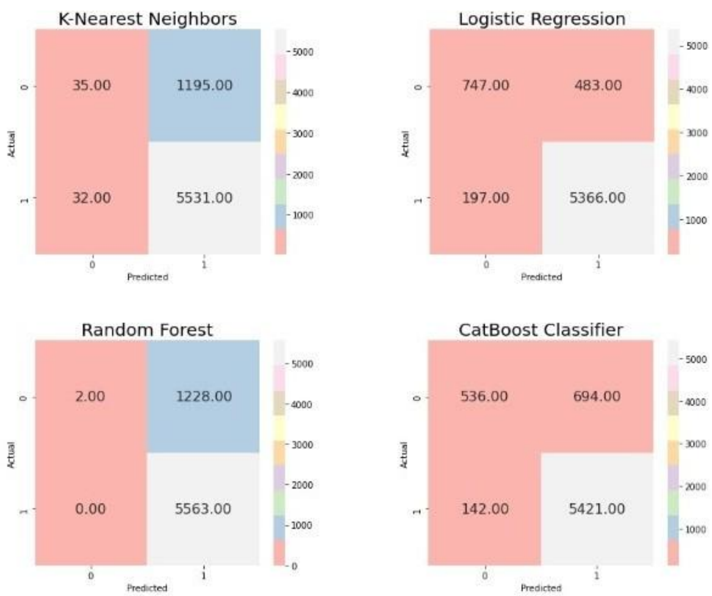


Fig. 13: Confusion matrices of the models used

Performance Indicators

In the aim of evaluating the performance of the applied models or the prediction of customers' product recommendation on the test set, we used different metrics, such as precision, recall, accuracy, and F1-score. They measure the ability of the predictive models to correctly predict the customer recommendation (0 or 1). The four indicators previously mentioned are calculated from the information captured using the confusion matrix. Where:

- The recall is the ratio of true churners or true positives (TP), and is calculated as follows:

$$R = TP/(TP+FN) \quad (1)$$

- The precision is the ratio of predicted correct churners, its formula is as follows:

$$P = TP/(TP+FP) \quad (2)$$

- The accuracy is the ration of the number of all correct predictions and is written as:

$$A = (TP+TN)/(TP+FP+TN+FN) \quad (3)$$

- The F-score is the harmonic mean of precision and recall and is written as follows:

$$F1 = (2 * Precision * Recall)/(Precision+Recall) \quad (4)$$

AUC

Ensemble Model

```
Accuracy: 0.9448753919788744
Classification Report:
              precision    recall  f1-score   support

           0       0.95      0.94      0.94      3039
           1       0.94      0.95      0.94      3020

    accuracy                           0.94      6059
   macro avg       0.94      0.94      0.94      6059
weighted avg       0.94      0.94      0.94      6059
```

AUC-ROC curve

To illustrate the diagnostic capability of binary classifiers, a Receiver Operator Characteristic (ROC) curve serves as a graphical representation. Summarizing each classifier's performance into a single measure might be helpful when comparing multiple classifiers. Calculating the AUC, often known as the area under the ROC curve, is one such strategy [18]. The AUC works well in real-world scenarios as a broad indicator of predictive accuracy. The AUC-ROC curve for the models utilized is depicted in the figure below (Fig. 14).

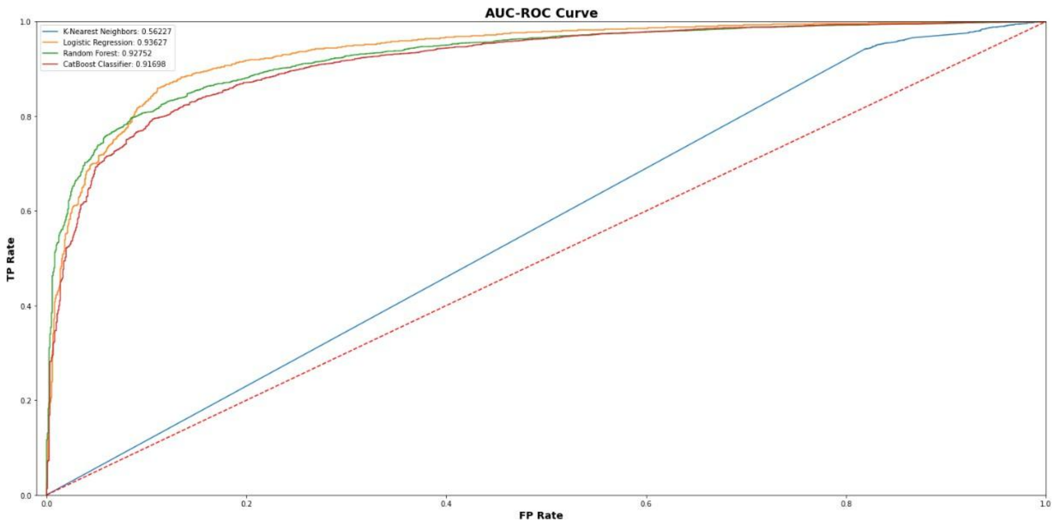


Fig. 14: AUC-ROC curve of the models used

Conclusion

The findings of this study underscore the potential of sentiment analysis in enhancing e-commerce strategies by providing insights into customer preferences. The Logistic Regression model emerged as the most effective tool for sentiment classification, achieving high accuracy and reliability. Future research should focus on improving model performance through advanced techniques such as deep learning and exploring real-time sentiment analysis applications. Additionally, addressing challenges such as the detection of fake reviews remains a critical area for further investigation .

Future Work

Future studies could explore the integration of deep learning techniques, such as recurrent neural networks (RNNs) and transformers, to further enhance sentiment analysis capabilities. Additionally, incorporating user interaction data and contextual information could lead to more nuanced sentiment understanding. Investigating the impact of sentiment analysis on actual purchasing behavior and customer retention would also provide valuable insights for e-commerce businesses.

References

[1] Kumar, V. V., Raghunath, K. M., Muthukumaran, V., Joseph, R. B., Beschi, I. S., & Uday, A. K. (2022). Aspect based sentiment analysis and smart classification in uncertain feedback pool. International Journal of System Assurance Engineering and Management, 13(1), 252-262.

[2] Wei, K., Li, Y., Zha, Y., & Ma, J. (2018). Trust, risk and transaction intention in consumer-to-consumer e-marketplaces: An empirical comparison between buyers' and sellers' perspectives. Industrial Management & Data Systems.

[3] Trivedi, S., & Patel, N. (2022). Mining Public Opinion about Hybrid Working with RoBERTa. Empirical Quests for Management Essences, 2(1), 31-44. Manal Loukili et al. 12

[4] Zad, S., Heidari, M., Jones, J. H., & Uzuner, O. (2021, May). A survey on concept level sentiment analysis techniques of textual data. In 2021 IEEE World AI IoT Congress (AIIoT) (pp. 0285-0291). IEEE.

[5] Ainin, S., Feizollah, A., Anuar, N. B., & Abdullah, N. A. (2020). Sentiment analyses of multilingual tweets on halal tourism. Tourism Management Perspectives, 34, 100658.

[6] Keikhosrokiani, P., & Pourya Asl, M. (Eds.). (2022). Handbook of Research on Opinion Mining and Text Analytics on Literary Works and Social Media. IGI Global.

[7] Chen, X., Xue, Y., Zhao, H., Lu, X., Hu, X., & Ma, Z. (2019). A novel feature extraction methodology for sentiment analysis of product reviews. Neural Computing and Applications, 31(10), 6625-6642.

[8] Mukherjee, S., & Bhattacharyya, P. (2012, March). Feature specific sentiment analysis for product reviews. In International conference on intelligent text processing and computational linguistics (pp. 475-487). Springer, Berlin, Heidelberg.

[9] Vayansky, I., & Kumar, S. A. (2020). A review of topic modeling methods. Information Systems, 94, 101582.

[10] Nagarajan, S. M., & Gandhi, U. D. (2019). Classifying streaming of Twitter data based on sentiment analysis using hybridization. Neural Computing and Applications, 31(5), 1425-1433.

[11] Fang, X., & Zhan, J. (2015). Sentiment analysis using product review data. Journal of Big Data, 2(1), 1-14.

[12] Yang, L., Li, Y., Wang, J., & Sherratt, R. S. (2020). Sentiment analysis for E commerce product reviews in Chinese based on sentiment lexicon and deep learning. IEEE access, 8, 23522-23530.

[13] Sindhu, C., & Mukherjee, D. (2021, April). A Joint Sentiment-Topic Model for Product Review Analysis of Electronic Goods. In 2021 5th International Conference on Computing Methodologies and Communication (ICCMC) (pp. 574 578). IEEE.

[14] Egger, R., & Gokce, E. (2022). Natural language processing (NLP): An introduction. In Applied Data Science in Tourism (pp. 307-334). Springer, Cham.

[15] Padmanabhan, B. (2020). Computational Personality Recognition and Sentiment Analysis of Select Novels of Cormac McCarthy. IUP Journal of English Studies, 15(3).

[16] Anandarajan, M., Hill, C., & Nolan, T. (2019). Text preprocessing. In Practical Text Analytics (pp. 45-59). Springer, Cham.

[17] Kabir, A. I., Ahmed, K., & Karim, R. (2020). Word Cloud and Sentiment Analysis of Amazon Earphones Reviews with R Programming Language. Informatica Economica, 24(4), 55-71. 13 Sentiment Analysis of Product Reviews for E-Commerce …

[18] Peng, J., Fung, J. S., Murtaza, M., Rahman, A., Walia, P., Obande, D., & Verma, A. R. (2022). A sentiment analysis of the Black Lives Matter movement using Twitter. STEM Fellowship Journal, (0), 1-11.